

# Proposed 2020 Census Data Products for Detailed Racial and Ethnic Groups

**Nicholas Jones**

**Director of Race Ethnicity Research & Outreach  
Population Division**

National Advisory Committee on Racial, Ethnic, and Other Populations – Fall 2021 Virtual Meeting  
November 4-5, 2021

Shape  
your future  
START HERE >

United States<sup>®</sup>  
**Census**  
**2020**

# Background

- Census Bureau has a long history of tabulating detailed race and ethnicity group data and protecting information provided by respondents
- Over time, more and more granular census data have been published
- Advances in data science, more powerful computers, and externally accessible 'big data' has increased the risk of identifying individuals from published statistics
- To mitigate, the Census Bureau is implementing new disclosure avoidance methods for the 2020 Census
- Differentially private Disclosure Avoidance Systems (DAS) are being developed to protect 2020 Census data products for detailed race and ethnicity groups, tribes, and their sex and age characteristics
- Census Bureau's goal is to meet data user needs while implementing new disclosure avoidance methods for 2020 Census data products

# Proposed 2020 Census Detailed DHC

- Population counts for detailed race, ethnicity, and tribal groups
- Basic demographic and housing characteristics for sex and age, household and family type, and tenure
  - Stakeholder feedback, combined with review of table access rates, and subject matter expert knowledge informed development of 2020 Census data products
  - 2020 Census data products include legally mandated statistics that will be supported by formal privacy systems

# Timeline for Developing Detailed-DHC

## Spring 2020

- DEMO team introduced to Group II Differential Privacy and Tumult Labs researchers
- Developed analytical tools for evaluating privacy and accuracy balance for producing detailed race ethnicity statistics

## Summer – Fall 2020

- DEMO researchers evaluated 2010 Census data and how best to achieve balance of privacy and accuracy

## Fall 2020

- DEMO presented proposal and recommendation to Census DSEP for total population and sex by age statistics
- DSEP provided initial approval for DEMO parameters and epsilon levels

## Winter – Summer 2021

- DEMO put Detailed DHC work on hold to review 2020 Census data files for producing apportionment and redistricting data products

# DEMO SafeTEx Analyses with Tumult Labs

Tumult Labs' SafeTEx tool allows Census Bureau researchers to assess interaction and balance of privacy and accuracy for producing detailed race, ethnicity, sex, and age data for the 2020 Census

Parameters for privacy and accuracy levels can be set to produce statistics on thousands of population groups

SafeTEx answers the question –

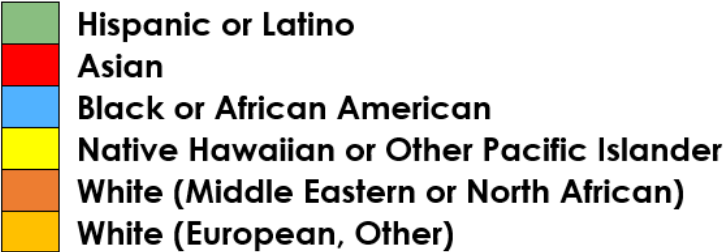
***Given specific parameter settings, what data could we publish within an acceptable level of accuracy?***

# 2020 Census Race and Ethnicity Code List

Appendix F.  
**Hispanic Origin and Race Code List**

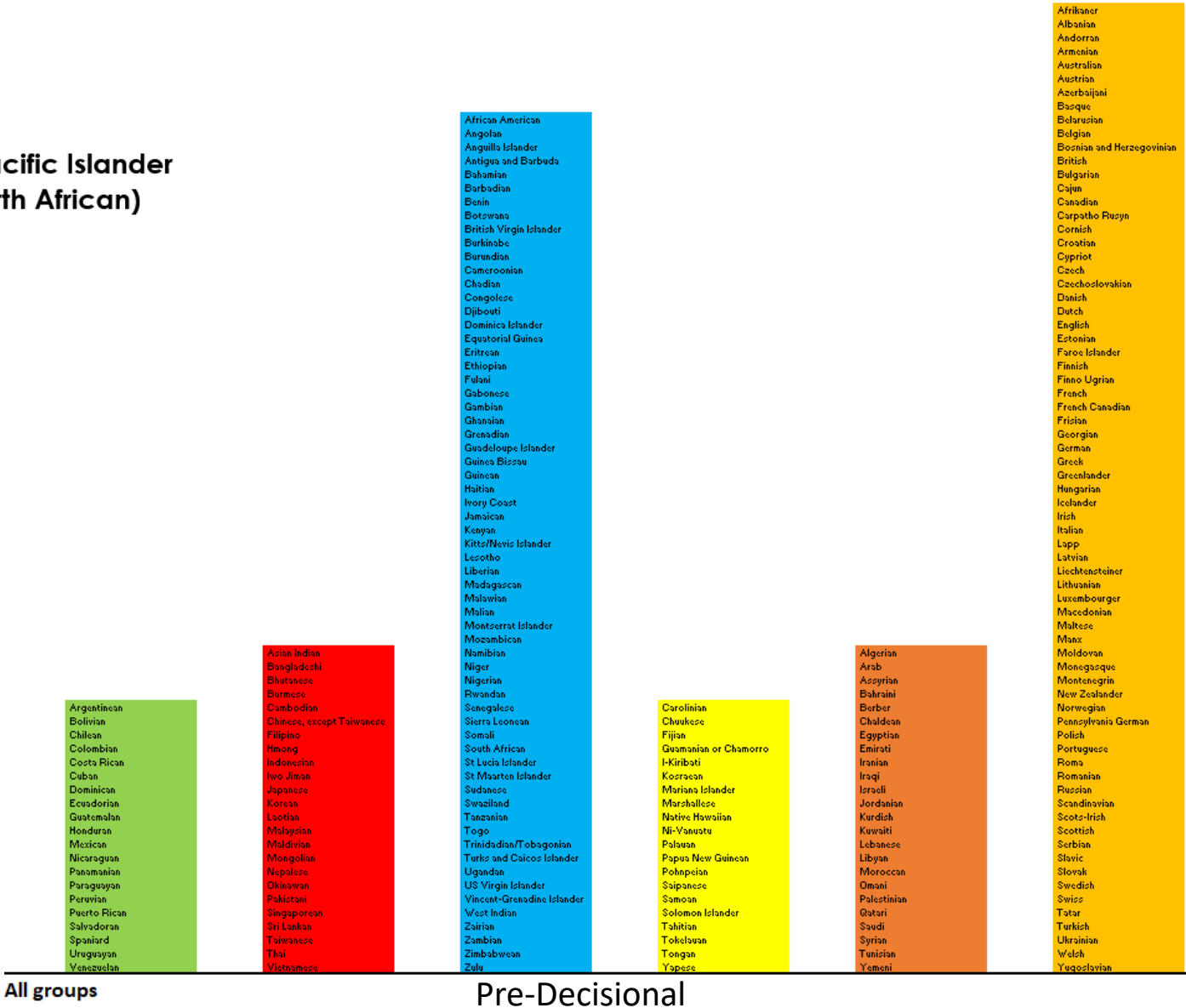
WHITE	1000-1999
<b>White</b>	<b>1000-1009</b>
White (Checkbox)	1000
White	1001
<b>EUROPEAN (EXCEPT SPANISH)</b>	<b>1010-1749</b>
<b>Albanian</b>	<b>1010-1014</b>
Albanian	1010
<b>Alsatian</b>	<b>1015-1019</b>
Alsatian	1015
<b>Andorran</b>	<b>1020-1024</b>
Andorran	1020
<b>Armenian</b>	<b>1025-1029</b>
Armenian	1025
<b>Austrian</b>	<b>1030-1034</b>
Austrian	1030
Tyrolean	1031
<b>Azerbaijani</b>	<b>1035-1039</b>
Azerbaijani	1035
<b>Basque</b>	<b>1040-1044</b>
Basque	1040
French Basque	1041

# Detailed Race and Ethnicity Groups



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY21-035.

\* Sources: 2010 Census SF-2 data, 2016 ACS 5-year data



Shape  
your future  
START HERE >

United States<sup>®</sup>  
Census  
2020



# Detailed American Indian Tribes and Alaska Native Village Groups

### American Indian tribes (federally recognized)

**American Indian** (general)

[illegible]

\* Source: 2010 Census CPH-T-6 data

### Alaska Native villages (federally recognized)

## Alaska Native (general)

[illegible]

ALL ALASKA NATIVE VILLAGES &amp; TRIBES



# Number of Detailed Groups UNDER Established Threshold at Various Levels

Threshold	White (n=95)	Black (n=63)	Asian (n=24)		NHPI (n=20)		AIAN (n=1,303)		Hispanic (n=20)
	Total	Total	Alone	AOIC	Alone	AOIC	Alone	AOIC	Total
100	0	0	2	1	2	1	490	453	0
200	0	1	2	2	6	2	613	573	0
300	1	4	2	2	8	3	690	652	0
400	3	5	2	2	8	4	753	710	0
500	3	5	2	2	10	6	793	754	0
600	6	6	2	2	11	7	833	786	0
700	6	10	2	2	11	7	867	817	0
800	6	13	2	2	11	7	894	836	0
900	6	13	2	2	11	7	910	860	0
1,000	6	14	2	2	12	9	930	880	0
1,500	10	20	2	2	13	11	981	945	0
2,000	11	22	2	2	13	11	1,017	990	0
2,500	12	23	2	2	13	12	1,053	1,018	0
3,000	13	26	3	2	14	12	1,080	1,049	0
3,500	13	26	4	2	15	12	1,099	1,064	0
4,000	13	27	4	2	15	12	1,112	1,080	0
4,500	13	30	4	2	15	13	1,120	1,104	0
5,000	15	30	4	2	16	13	1,133	1,113	0

	0 groups
	1-15 groups
	16-30 groups
	More than 100 groups

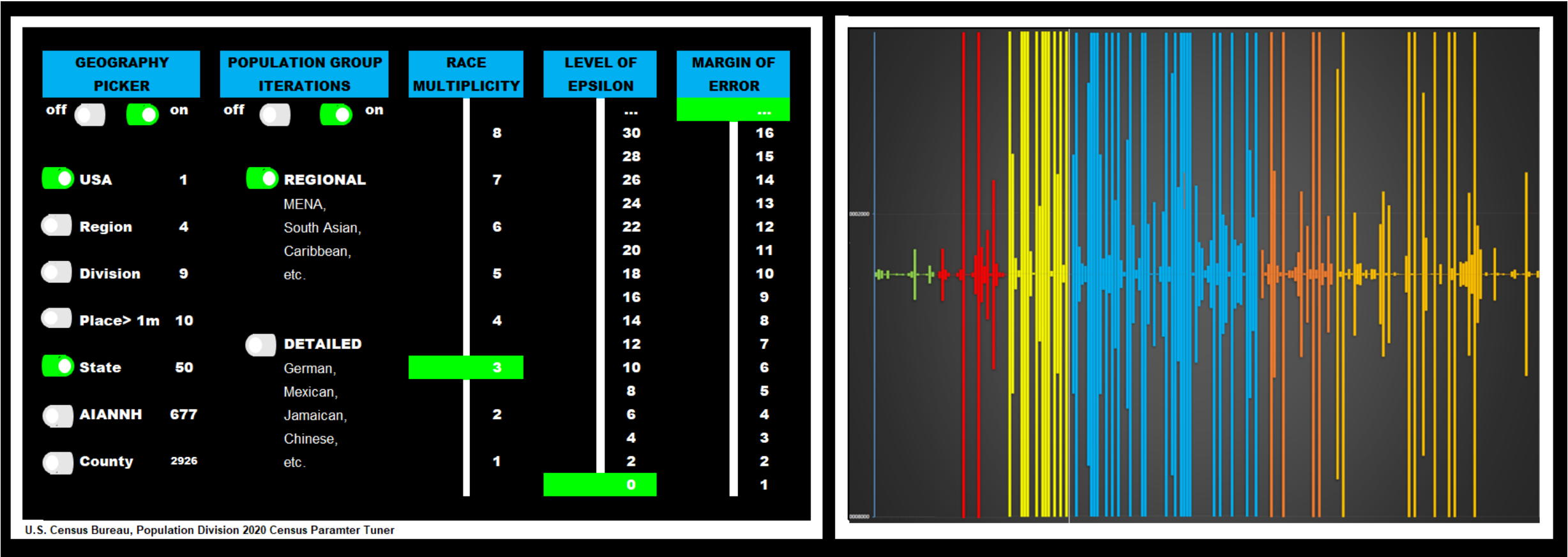
\* Sources: 2010 Census Asian Brief, 2010 Census Pacific Islander Brief, 2010 Census SF-2 data, 2010 Census CPH-T-6, 2016 ACS 5-year data

# DEMO Research Goals

**Develop parameter settings for algorithm to produce differentially private 2020 Census data for detailed race and ethnicity groups, and their sex and age characteristics**

- Explore how to establish privacy and accuracy levels to produce statistics for our nation's myriad detailed racial and ethnic population groups
- Allow for equity across all major race and ethnicity categories
- Provide same amount, or more data, than provided from 2010 Census

# Parameter Tuner and Noise Visualization



- Hispanic or Latino
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White (Middle Eastern or North African)
- White (European, Other)

# 2020 Census Detailed DHC

## Table 1: Population Counts

Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells					
2020 Table #	Cell Count	Indent	Table Shell		
T1.		0	TOTAL POPULATION [1]		
T1.		0	Universe: Total population		

# 2020 Census Detailed DHC

## Table 2: Sex by Age

Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells						
2020 Table #	Cell Count	Indent	Table Shell			
T2.			0 SEX BY AGE FOR SELECTED AGE CATEGORIES [49]			
T2.			0 Universe: Total population			
T2.	1	0	Total:			
T2.	2	1	Male:			
T2.	3	2		Under 5 years		
T2.	4	2		5 to 9 years		
T2.	5	2		10 to 14 years		
T2.	6	2		15 to 17 years		
T2.	7	2		18 and 19 years		
T2.	8	2		20 years		
T2.	9	2		21 years		
T2.	10	2		22 to 24 years		
T2.	11	2		25 to 29 years		
T2.	12	2		30 to 34 years		
T2.	13	2		35 to 39 years		
T2.	14	2		40 to 44 years		
T2.	15	2		45 to 49 years		
T2.	16	2		50 to 54 years		
T2.	17	2		55 to 59 years		

# 2020 Census Detailed DHC

## Table 3: Household Type

Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells						
2020 Table #	Cell Count	Indent	Table Shell			
T3.		0	HOUSEHOLD TYPE [9]			
T3.		0	Universe: Households			
T3.	1	0	Total:			
T3.	2	1	Family households:			
T3.	3	2	Married couple family			
T3.	4	2	Other family:			
T3.	5	3	Male householder, no spouse present			
T3.	6	3	Female householder, no spouse present			
T3.	7	1	Nonfamily households:			
T3.	8	2	Householder living alone			
T3.	9	2	Householder not living alone			
			NOTE: When a category other than Total Population is selected, the data in this table refer to the race, Hispanic or Latino origin, or tribe of the householder.			

# 2020 Census Detailed DHC

## Table 4: Tenure

Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells						
2020 Table #	Cell Count	Indent	Table Shell			
T4.		0	TENURE [4]			
T4.		0	Universe: Occupied housing units			
T4.	1	0	Total:			
T4.	2	1	Owned with a mortgage or a loan			
T4.	3	1	Owned free and clear			
T4.	4	1	Renter occupied			
NOTE: When a category other than Total Population is selected, the data in this table refer to the race, Hispanic or Latino origin, or tribe of the householder.						



# Data Product Comparisons: 2010 vs 2020

Topics	2010	2020
Number of Population Tables	57	4
Number of Geographies	104	4
Threshold for National Sex by Age Characteristics	100	500
Threshold for National Total Count for <ul style="list-style-type: none"> <li>• AIAN Tribes and Villages</li> <li>• Detailed Race and Ethnicity Groups</li> </ul>	0 100	10 10
Number of Race and Ethnicity Groups (not inclusive of AIAN tribes and villages)	67	334
Disclosure Avoidance Technique	Swapping	Differential Privacy

Note: Estimates based on 2010 Census SF-2 data, 2016 ACS 5-Year data, and 2020 Census proposal for Group II data products.

# Next Steps for Developing Detailed DHC

## Where Are We Now?

### Fall 2021...

- DEMO researchers re-engaging with R&M and Tumult after completion of 2010 Census PL release
- Engaging with data users and public on Detailed DHC proposal and use cases
- Evaluate privacy/accuracy tradeoff for the Detailed DHC data

### Before “final” parameter setting and epsilon decisions are made...

- Discuss proposed 2020 Census Group II data product plans with tribal leaders in formal AIAN Tribal Consultations
- Address feedback and recommendations from Census Bureau Advisory Committees
- Assess implications that post-processing will have on production of detailed race and ethnicity data products

# Questions for NAC

**What questions or concerns do you have about any potential changes to the DHC and/or the Detailed DHC?**

**Are there any critical decennial census statistics that are not currently proposed in the 2020 Data Products crosswalk and/or geographic detail that you require?**

**What levels of geography (e.g., places, school districts) are needed for detailed race ethnicity population groups, and for what purpose, policy, program, etc.?**